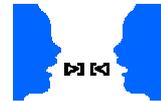


CHILDES

<http://childes.psy.cmu.edu>

CHILDES: Child Language Data Exchange System



The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video.

System

[Ground rules](#)

[Membership list](#)

How to subscribe to [info-CHILDES](#)

[Overviews](#)

Programs and Data

[CLAN programs](#)

[CHILDES Database](#)

[XML Database](#)

[Contributing new data](#)

Links

[Child Language sites](#)

The [TalkBank](#) Project

The [LIDES](#) Database

Manuals (in PDF)

[CHAT transcription system manual](#)

[BTS sign transcription manual](#)

[CLAN program manual](#)

[CHILDES database manual](#)

Teaching with CHILDES

[Topics](#) in language acquisition.

[On-line tutorial](#) for learning CHAT and CLAN

[Tips](#) from teachers.

Special Tools

Language acquisition [bibliography](#)

[UniCode, MOR, POST, CMULEX](#)

Parental frequency [count](#) and [documentation](#).

Procedures

Procedures and tools for [CA analysis](#)

[Slicing and splicing](#) audio files with Praat.

Working with [digitized video](#) | [digitized audio](#)

Contacts

[Pittsburgh](#), USA - [Brian MacWhinney](#)

[Antwerp](#), Belgium - [Steven Gillis](#)

[Chukyo](#), Japan - [Hidetosi Sirai](#)

Antwerp webmaster: [Joris Gillis](#)

Last modified: 08/19/2003 13:41:23



CLAN files

The header

@Begin
@Languages: en
@Participants: CHI Peter Target_Child, MOT Mother, PAT Patsy Investigator,
LOI Lois Investigator
@ID: en|bloom70|CHI|1;9.8|male|normal||Target_Child||
@ID: en|bloom70|MOT|||||Mother||
@ID: en|bloom70|PAT|||||Investigator||
@ID: en|bloom70|LOI|||||Investigator||
@Date: 4-OCT-1971
@Comment: MLU 1.04 the majority of it. Morpheme utterances were imitations of an immediately preceding utterance, spontaneous 2 morpheme utterances included "blocks" and "fix it" and "go down".
@Situation: Playing. The interaction took place almost entirely in the living room which has a large picture window overlooking Morningside Park.
@Activities: Peter crawling, pretending to swim, down hall to living room where Patsy and Lois wait
@Comment: Peter was mostly interested in his own blocks and in his book "Pat the Bunny," the slides, the see-saw and the tape recorder. He also enjoyed running to the window and standing on the sill to look for things he could hear like fire engines and dogs. Peter's mother was at home and not in the room during the sessions. Peter gets along well without her presence. He paid no attention to Jennifer except to comment when she cried.

Speaker and dependent tiers

*CHI: xxx this is another .
%mor: pro:dem|this v|be&3S det|another .
%act: taking out more train cars

*CHI: xxx this is a train too .
%mor: pro:dem|this v|be&3S det|a n|train adv|too .

*CHI: you got any # train tracks .
%mor: pro|you v|get&PAST qn|any n|train n|track-PL .
%int: <1w-5w> rising

*LOI: no .
%mor: co|no .

Special characters

Word symbols

xxx	unintelligible speech, not treated as a word
xx	unintelligible speech, treated as a word
yyy	unintelligible speech, transcribed on %pho line, not treated as a word
yy	unintelligible speech, transcribed on %pho line, treated as a word
www	untranscribed material
&	phonological fragment
()	incomplete word
0word	omitted word

Morpheme symbols

-	suffix marker
#	prefix marker
+	compound boundary
&	fusion

Utterances terminators

.	period
?	question
!	exclamation
-?	rising final contour
-.	falling final contour
#	pause between words
-:	previous word lengthened

Special terminators and linkers

+...	trailing off
+/.	interruption
+,	self completion
++	other completion

Prosody within words

/	stress
//	accented nucleus
:	lengthened syllable
::	pause between syllables

Scoped symbols

[= pushing car]	paralinguistic
[% looks sleepy]	comment on main line
[!]	stressing
[/]	retracing
[//]	retracing with correction

CLAN

Start CLAN

- To start CLAN press on the CLAN icon (or a CLAN file) and then press the COMMANDS button

Setting directory:

- press 'working' > select files in window > press 'select current directory'
- press 'lib' > select files > press 'select current directory'
- the output directory is the same as the working directory unless you specify a different directory

Typing in a command:

- There are two ways to build up commands: (1) you can type them in, (2) you can access a list of prespecified commands when you press the CLAN button
- Once you have used a specific command you can use this command again (1) if you press the RECALL button, which gives you a list of the last 20 commands, and double-click on the command you need, (2) if you press ↑, which enters the previous command into the COMMAND window.

Selecting files:

- Before you run a command you have to select files from the directory: Press the FILE IN button and move the files you wish to analyse to the window 'Files for Analysis'. After the files are selected and you have returned to the Commands window, an @ is appended onto the command string, which represents the set of files listed.

COMBO

COMBO provides the user with ways of composing Boolean search strings to match patterns of letters, words or groups of words in the data files. This program is particularly useful for doing syntactic analyses. The search strings are specified with either the +s/-s option or in a separate file. Use of the s-switch is obligatory in COMBO.

Boolean operators

^	AND	COMBO searches for a specified sequence of words: +s"big^dog"
+	inclusive OR	COMBO searches for at least one of two words: +s"cat+dog"
*	wildcard	COMBO searches for everything: +s"walk*"
+x	order	If you add this switch to an 'and' command (e.g. +s"big^dog"), COMBO will search for 'big' and 'dog' in both orders.

The Boolean operators can be combined: +s"big*^*dog"

Using search files

The +s switch can also be used to search for items specified in a file. Type in +s@filename:

+s@prepositions

You need to create a file called 'prepositions' which includes a list of prepositions, one on each line. COMBO will use this file to search the data for all utterances including one of the prepositions included in the search file. Search files can be combined:

+s@prepositions^@articles

In this case you need two search files including a list of prepositions and a list of articles. The output file will include all utterances in which prepositions are followed by an article.

General CLAN commands

Input file

Each command must include the name of the input file. If several files are include in the input, use the * symbol.

Combo *s”dog” Peter*

Output file

By default CLAN sends the output to the screen. If you wish to create an output file you have several options:

+f It allows you to send the output to a file rather than the screen. The output file will be in the same directory as the working file. If you search several files, +f will create several output files.

> In order to send the out put from all files to a single output file use the arrow > and specify the name of the output file.

... > prepositions.cha

>> The single arrow > overwrites material already in the file. If you want to append material that is already in the output file use a double arrow >>

... >> prepositions.cha

The single arrow and an ampersand >& gives you important system messages in addition to the output.

+u It provides you with another opportunity of directing the output of several search files into a single output file. Note that if too many files are selected, CLAN is unable to merge them with this command.

Context

+w / -w These options give you the opportunity to include context into the output. The switches work only with COMBO and KWAL. The +w plus number indicates the number of tiers preceding the target tier that should be included in the output; the -w plus number indicates the number of tiers following the target tier that should be included in the output.

... +w3 -w3 > ...

File information in the output

The output may or may not include information about the location of the target utterance in the source file. You have four options

- Ø If you type in nothing, COMBO indicates the location of the tier where it occurs.
- +d If you add +d, COMBO omits all information about the source file.
- +d1 If you add +d1, COMBO indicates line numbers and file names.
- +d2 If you add +d2, COMBO indicates file names only once per file.

Tiers

- +t / -t It allows you to include or exclude particular tiers. For instance, if you request to analyze the speaker *MOT and all the %cod dependent tiers, the program will analyze all of the *MOT main tiers and only the %cod tiers associated with that speaker.
- +o This switch allows you to specify tiers that should be exempted from the +t/-t switch command

Pipes

- | It sends the output from one command through a pipe to a second command. For instance, you can send to output of a COMBO command to FREQ.

```
combo +s"the" +t*CHI sample.cha +d | freq
```

The switches can occur in any order in the command window with one exception: the CLAN commands (e.g. COMBO, FREQ) must come first.

FREQ

FREQ produces a list of all words used in the file along with their frequency counts and calculations of type-token ratio. The type-token ratio is found by calculating the total number of unique words used by a selected speaker and dividing that number by the total number of words used by the same speaker. It is generally used as a rough measure of lexical diversity. Of course, the type-token ratio can only be used to compare samples of equivalent size because as sample size increases the increase in the number of types starts to level off.

FREQ ignores the following strings in the data:

1. xxx (unintelligible words)
2. www (untranscribable words for technical reasons)
3. words beginning with one of the following characters: 0, &, +, -, #

The +s switch gives you the opportunity to determine the frequency of specific words. You can also use a search file to determine the frequency of a prespecified list of words. Moreover, the +t switch gives you the opportunity to limit your search to a specific speaker.

```
freq +s@demonstratives *CHI sample.cha
```

If you don't indicate anything, freq will count all words of all speakers in all input files and will present them in alphabetical order. You can use the +o switch to create a frequency list in descending frequency order, and you can use the +o1 switch to create a frequency list in ascending frequency order

MLU

The MLU command computes the mean length of utterances, measured by the average number of morphemes in an utterance. The MLU is commonly seen as a measure of grammatical development. However, the measure is only reliable until about 48 months, or an MLU of 3.0.

In order to use the MLU you have to insert morpheme boundaries on the speaker tiers or you use the %mor-tier. If you use the %mor-tier you have to indicate this tier in the command line.

```
mdu +t%mor sample.cha
```

However, you can also run the MLU command on the main tier without specified morpheme boundaries. In this case, the MLU program computes the average number of words per utterance, rather than the average number of morphemes. However, in English the average number of words per utterance correlates with very closely with the average number of morphemes per utterance (at a level of about 97%).

Many researchers use only 100 child utterances to determine the MLU. CLAN provides an opportunity to specify a search domain of only 100 utterances. However, it is easier to use a complete file.

The MLU command does not consider the following strings: xxx, yyy, www, uh um, 0* &* +* -* #* \$*. Also excluded is material in [/] brackets.